DOCUMENT RESUME

ED 253 566 TM 850 106

AUTHOR Stallings, William M.
TITLE Mind Your p's and Alphas.
PUB DATE 85
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Research; *Hypothesis Testing; Multivariate Analysis; *Probability; *Research Methodology; Research Problems; *Statistical Significance; Statistical Studies
IDENTIFIERS *Alpha Coefficient

ABSTRACT

In the educational research literature alpha, the a priori level of significance, and p, the a posteriori probability of obtaining a test statistic of at least a certain value when the null hypothesis is true, are often confused. Explanations for this confusion are offered. Paradoxically, alpha retains a prominent place in textbook discussions of such topics as statistical hypothesis testing, multivariate analysis, power, and multiple comparisons while it seems to have been supplanted by p in current journal articles. The unique contributions of both alpha and p are discussed and a plea is made for using both conventions in reporting empirical studies. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED253566

Mind Your p's and Alphas

William M. Stallings

Georgia State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. M. Stallings

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 820 106

Running head: Mind Your p's and Alphas

Abstract

In the educational research literature alpha, the a priori level of significance, and p, the a posteriori probability of obtaining a test statistic of at least a certain value when the null hypothesis is true, are often confused. Explanations for this confusion are offered. Paradoxically, alpha retains a prominent place in textbook discussions of such topics as statistical hypothesis testing, multivariate analysis, power, and multiple comparisons while it seems to have been supplanted by p in current journal articles. The unique contributions of both alpha and p are discussed and a plea is made for using both conventions in reporting empirical studies.

Mind Your p's and Alphas

When teaching introductory statistics or research methods, I find that my students invariably have difficulty distinguishing alpha, the a priori level of significance, from p, the a posteriori probability of obtaining a test statistic of at least a certain magnitude when the null hypothesis is true. I sympathize with my students for there is indeed a dissonance between what is presented in the textbooks and what is found in empirical research.

Introductory texts tend to present statistical hypothesis testing in a decision-theoretic mode (but see Hays, 1981). Alpha is described as a "ground rule" needed to be set by the researcher before gathering data or, in another conceptualization, before calculating the test statistic. In these texts p receives little attention -- except when mechanically compared with alpha for determining whether to reject the null hypothesis. In published empirical research the relative importance of alpha and p seems to be reversed. That is, more emphasis is placed on such summaries as t (22) = 2.62, $p < .01$ than on verbal statements to the effect that the null hypothesis was rejected at the .01 level of significance. Do the two conventions, alpha and p, convey the same information? I tend to agree with Minium's (1978) position.

> Some researchers evaluate the outcome of the test of a hypothesis by showing the probability of obtaining a value as discrepant as the one obtained if the null hypothesis were true. For a given outcome they might

> report, say, that "$p < .03$." Now this probability statement is an expression of the rarity of the sample outcome and nothing more. It cannot be interpreted as the value of [alpha] .... these persons who employ both = .05 and .01 to evaluate the several tests they have conducted may be using these values in the sense of decision criteria or as a rough statement of rarity of sample outcomes .... one often suspects that the authors themselves have not thought through exactly what they intend. (p. 277)

Of course, one can allocate differentially the overall alpha for a set of contrasts by using the Bonferroni inequality (Kirk, 1982, p. 109). Still, Minium's point is well taken.

Despite such criticisms, it is obvious to anyone who reads the research literature that the rigid distinction between p and alpha has broken down and that there is a decreasing tendency for authors to provide explicit alpha values while, simultaneously, there is an increasing trend of reporting p values only. We need to confront afresh the meanings and usages of p and alpha, both singly and in conjunction.

Let's begin with basics. Although alpha and p are both probabilities, alpha is a constant while p is a random variable. p is sensitive to sample size; alpha is not.

To determine the possibility of making type I and type II errors alpha and p are needed. p, when compared to a particular

alpha, may be interpreted in terms of type II error (not beta but to the possibility of making a type II error). For example, if p exceeds alpha, the usual decision is not to reject the null hypothesis, thus incurring the risk of a type II error. On the other hand, if $p <$ alpha, the usual decision is to reject the null hypothesis and there is no resultant possibility of making a type II error (obviously a type I error may have been committed).

Although alpha can be chosen on rational grounds, alpha tends to be selected by tradition. The usual values are .05 and .01 and more rarely .1 and .001. Parenthetically, the tradition itself may be rational (Cowles & Davis, 1982). However selected, alpha values are a necessary component for estimating a priori power and for determining sample size. Additionally, the control of type I error is a primary consideration in using multivariate analysis (Strahan, 1982) and selecting multiple comparison techniques.

According to the most recent edition of the Publication Manual of the American Psychological Association (APA, 1983) in the "Abstract" section of empirical studies one should describe "the findings, including statistical significant levels" (p. 24). Should researchers use alpha or p or both as decision rules? Or should they be allowed to have it both ways? [For example, to claim that $p < .06$ is close enough to alpha = .05 to be called a statistically significant finding; this seems to be Tatsuoka's

(1982, pp. 1781-1782) position]. Given that researchers are called upon to highlight statistical significant levels, it is unfortunate that the Publication Manual (APA, 1983) gives as illustrations, $p < .05$ and $p < .01$, illustrations which suggest a confusion or confounding of alpha with p.

Perhaps, the confusion between alpha and p is historical. Operating under his fiducial theory of statistical inference (hence, the terms may be, in Kuhn's conception, "incommensurable"), Fisher (1947, p. 13) seemed to use level of significant as a synonym for p and standard (italics mine) level of significance for what we call alpha.

In the empirical literature, as well as in the Publication Manual, p's are often reported in values associated with alpha (e.g., .05 and .01). There are, at least, two explanations: (a) statistical tables (such as those for F) often give only .05 and .01 probabilities and (b) since "exact" probabilities are obtainable only if the null hypothesis is true and if the mathematical model is correct, the researcher may feel that "ball park" numbers provide sufficient accuracy. Whatever the explanation, there is no compelling reason not to either report "exact" probabilities (easily done with computers) or probability intervals (e.g., $.20 > p > .10$).

Of course, p values are informative. They permit the reader to impose his/her own alpha level on reported findings (Hopkins & Glass, 1978, p. 243). Moreover, p's are necessary for combining

probabilities and conducting some forms of meta-analysis. And even after a "rejection at the .05 level of significance," a *p* value of say .01 gives "greater assurance than the .05 level that H_o is false" (Hopkins & Glass, 1978, p. 223; see also Kerlinger, 1979, p. 314). Similar, but more dramatic, is Lehmann's (1968) contention that *p* is an "index of surprise." As Lehmann elaborated, "the smaller [our *p*] is, the more surprising it is to get this extreme a value under H [the null] and, therefore, the stronger the evidence against H" (p. 43). According to Tatsuoka (1971) the discrepancy between *p* and alpha for Hotellings T^2 and MANOVA gives indirect evidence about possible violation to the underlying mathematical assumptions.

> If the results are borderline significant, one should be cautious of the conclusions unless there is good evidence that the assumptions are not badly violated; if the results are significant far beyond the pre-assigned level, then one need not be concerned about whether the assumptions have been met. (p. 51)

To repeat, *p* values are informative, but they are even more informative when coupled with an explicit alpha.

As we all know, *p* is now serving both as alpha and as *p*, a role starkly described by Lackritz (1984): "Investigators have come to rely on *p* values in making decisions about hypotheses ...." (p. 312). Despite current usage, alpha and *p* are different; together, they can provide non-redundant information to the

researcher and the reader of research. This non-redundancy is most evident if exact p's or p intervals are given. An empirical article will be more informative if it reports that alpha was set at .05 and that $p = .031$. Perhaps we should adopt a new convention, for example, one illustrated by $p = .031 < .05$ where the alpha level would follow the inequality sign.

References

American Psychological Association. (1983). Publication manual of the American Psychological Association (3rd ed.). Washington, D.C.: Author.

Cowles, M. & Davis, C. (1982). On the origins of the .05 level of statistical significance. American Psychologist, 37 (5), 553-558.

Fisher, R. A. (1947). The design of experiments (4th ed.). New York: Hafner.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt Rinehart and Winston.

Hopkins, K. D. & Glass, G. V (1978). Basic statistics for the behavioral sciences. Englewood Cliffs, New Jersey: Prentice-Hall.

Kerlinger, F. N. (1979). Behavioral research: A conceptual approach. New York: Holt, Rinehart and Winston.

Kirk, R. E. (1982). Experimental design (2nd ed.). Belmont, California: Brooks/Cole.

Lackritz, J. R. (1984). Exact p values for F and t tests. American Statistician, 38 (4), 312-314.

Lehmann, E. L. (1968). Hypothesis testing. In D. L. Sills (Ed.), International Encyclopedia of the Social Sciences: Vol. 7 (p. 40-47). New York: Macmillan and Free Press.

Minium, E. W. (1977). Statistical reasoning in psychology and education (2nd ed.). New York: Wiley.

Straham, R. F. (1982). Multivariate analysis and the problem of type I error. Journal of Counseling Psychology, 29 (2), 175-179.

Tatsuoka, M. M. (1982). Statistical methods. In H. E. Mitzel (Ed.), Encyclopedia of Educational Research: Vol. 4 (pp. 1780-1808). New York: Free Press.

Tatsuoka, M. M. (1971). Significance tests: Univariate and multivariate. Champaign, Illinois: Institute for Personality and Ability Testing.